



Marco filogenómico e historia de la divergencia de los Cephalochordata Amphioxus

Qi-Lin Zhang^(1,2†), Guan-Ling Zhang^{1†}, Ming-Long Yuan^{3†}, Zhi-Xiang Dong¹, Hong-Wei Li¹, Jun Guo¹, Feng Wang¹, Xian-Yu Deng¹, Jun-Yuan Chen^{2,4*} y Lian-Bing Lin^{1*}.

¹ Facultad de Ciencia y Tecnología de la Vida, Universidad de Ciencia y Tecnología de Kunming, Kunming, China, ² Instituto Evo-Devo, Facultad de Ciencias de la Vida, Universidad de Nanjing, Nanjing, China, ³ Laboratorio Estatal Clave de Agroecosistemas de Pastizales, College of Pastoral Agricultural Science and Technology, Lanzhou University, Lanzhou, China, ⁴ State Key Laboratory de Paleobiología y Estratigrafía (LPS), Instituto de Geología y Paleontología de Nanjing, Academia China de Ciencias, Nanjing, China

OPEN ACCESS

Editado por:

Ruriko Yoshida,
Universidad de Kentucky, Estados Unidos

Revisado por:

Neil Moore,
Universidad de Kentucky, Estados Unidos
Grady Weyenberg,
Universidad de Hawai en Hilo,
Estados Unidos
Joseph Rusinko,
Hobart and William Smith Colleges,
Estados Unidos

*Correspondencia:

Jun-Yuan
chenjy@nju.edu.cn
Lian-Bing
linlb@kmust.edu.cn

(†) Estos autores han contribuido
igualmente a este trabajo

Sección especializada:

Este artículo se envió a
Biología
de
sistemas, una sección de
la revista Frontiers in
Physiology

Recibido: 08 de septiembre de 2018

Aceptado: 06 de diciembre de 2018

Publicado: 18 de diciembre de 2018

Cita:

Zhang Q-L, Zhang G-L,
Yuan M-L, Dong Z-X, Li H-W, Guo J,
Wang F, Deng X-Y, Chen J-Y y Lin L-B
(2018) Un marco filogenómico y la historia
de divergencia de Cephalochordata
Amphioxus.
Front. Physiol. 9:1833.
doi: 10.3389/fphys.2018.01833

Los anfioxos, o cefalocordados, se utilizan a menudo como el sustituto invertebrado viviente de los ancestros vertebrados y son ampliamente utilizados como modelos de biología evolutiva de los cordados. Sin embargo, su filogenia, historia de divergencia y características de especiación siguen siendo poco conocidas, y los estudios filogenómicos para explorar estos problemas faltan por completo en la literatura. Aquí, hemos determinado un nuevo transcriptoma de *Branchiostoma japonicum*. Combinado con secuencias masivas de las otras 18 especies, se construyó una filogenia de 19 vías a través de múltiples métodos (ML, BI, PhyloBayes y ASTRAL), apoyando consistentemente una filogenia de [(*B. belcheri*+ *B. japonicum*)+ (*B. lanceolatum*+ *B. floridae*)+ *Asymmetron lucayanum*] en anfioxo. Se encontraron señales filogenéticas congruentes entre los genes mitocondriales, el ARN 12S y los genomas mitocondriales completos según informes anteriores, lo que indica que el ARN 12S puede tener potencial como marcador molecular para el análisis filogenético en los anfioxos. El análisis de datación molecular indicó una radiación de los cefalocordados durante el Cretácico (~ hace 104-61 millones de años), apoyando una asociación entre la diversificación y especiación de los cefalocordados con la deriva continental y los cambios asociados en sus respectivos hábitats durante este tiempo. El análisis de enriquecimiento funcional identificado para los dominios específicos de cada especie indicó que su función implica principalmente la respuesta inmunitaria, la apoptosis y el metabolismo y la utilización de lípidos, lo que indica que los patógenos y los cambios en las necesidades energéticas son una importante fuerza impulsora de la especiación de los anfioxos. Este estudio representa el primer análisis filogenómico a gran escala de la mayoría de los principales géneros de anfioxos basado en datos filogenómicos, proporcionando una nueva perspectiva tanto de la filogenia como de la especiación divergente de los cefalocordados.

Palabras clave: Cephalochordata, amphioxus, filogenómica, historia de divergencia, proceso de especiación, dominios proteicos, genómica evolutiva.

INTRODUCCIÓN

Amphioxus, también conocido como lancetas, pertenece a Leptocardii: Amphioxiformes y es el único representante moderno del subfilo Cephalochordata (Koop y Holland, 2008). Los Amphioxus son un tema clave para las investigaciones de biología evolutiva, ya que aportan indicios sobre los orígenes evolutivos de los vertebrados y constituyen un intrigante punto de comparación para rastrear cómo

vertebrados han evolucionado y se han adaptado (Shimeld y Holland, 2005; Putnam et al., 2008). A pesar de la evolución separada de cefalocordados y vertebrados a partir de un ancestro común hace más de 520 millones de años (Mya), su estructura corporal y morfología conservaron al máximo características de estos ancestros vertebrados (*Haikouella lanceolata*) (Chen et al., 1999). Con la reciente acumulación de datos genéticos sobre los anfioxus, la filogenia molecular, la biología evolutiva del desarrollo (evo-devo) y la genómica sugieren además que los anfioxus son un modelo válido de ancestros vertebrados (Shimeld y Holland, 2005; Putnam et al., 2008). El anfioxo actual es el único cefalocordado viviente y se distribuye en tres géneros: *Branchiostoma* (~ 28 especies) con el mayor número, *Epigonichthys* (una especie), y *Asymmetron* (una especie). Sin embargo, podrían existir especies crípticas adicionales (Yue et al., 2014). *Branchiostoma* está ampliamente distribuido en las latitudes medias-bajas del océano Atlántico, el Mediterráneo y el océano Pacífico, y los dos géneros restantes se distribuyen principalmente en las latitudes medias-bajas y los trópicos a lo largo de la zona ecuatorial. El hábitat de los anfioxus se encuentra en aguas marinas costeras poco profundas, entre 8 y 16 m, y los animales esconden la mayor parte de su cuerpo en la arena fina del fondo marino, encontrando su alimento mediante la alimentación por filtración del agua de mar.

A pesar del importante papel de los anfioxus en la biología evolutiva, las publicaciones que investigan la filogenia molecular dentro de los anfioxus son escasas, especialmente en los últimos . Por el contrario, muchos estudios se centran en cuestiones científicas relacionadas con la biología evolutiva y la filogenia molecular entre anfioxos y vertebrados, lo que conduce a un sesgo en nuestro conocimiento de la evolución de los anfioxos. Los estudios previos de la última década se centraron en la taxonomía de los amphioxus y la divergencia entre especies basándose en la genómica mitocondrial (ADN mitocondrial, ADNmt). Por ejemplo, Luo et al. (2007) utilizaron la subunidad 1 de la citocromo c oxidasa (*cox1*) y fragmentos del gen de la subunidad b del citocromo (*cob*) para construir árboles filogenéticos entre *Branchiostoma belcheri*, *B. japonicum* y *B. tsingtaoensis*, y se calculó la distancia genética media entre ellos. Los resultados mostraron un clado agrupado y diferencias intraespecíficas entre *B. japonicum* y *B. tsingtaoensis*, apoyando la clasificación de *B. tsingtaoensis* como *B. japonicum*. Zhong et al. (2009) secuenciaron el ADNmt completo de *B. japonicum* y *B. belcheri* y combinaron los resultados con los 13 genes codificadores de proteínas GCP del ADNmt de otras cinco especies de anfioxus para construir un árbol de máxima verosimilitud (ML) de secuencias de proteínas y ácidos nucleicos, respectivamente. Ambos árboles filogenéticos apoyaron los resultados de Luo et al. (2007); además, los resultados de estas investigaciones indicaron que los anfioxus que viven en Beihai, provincia de Guangxi, y Maoming, provincia de Guangdong, pertenecen a la misma especie *B. belcheri* (Zhong et al., 2009). Además, Zhong et al. (2009) construyeron un árbol filogenético con diferentes poblaciones geográficas de anfioxus basado en la secuencia 12S rRNA (ARN ribosómico) del ADNmt. Cabe señalar que el árbol construido por 13 PCGs apoyó la siguiente estructura topológica: (((*B. belcheri*+ *B. japonicum*)+ *Branchiostoma lanceolatum*)+ *Branchiostoma floridae*)+ *Asymmetron*), mientras que el árbol basado en el ARNr 12S no apoyó *B. lanceolatum*

como linaje hermano del clado (*B. belcheri*+ *B. japonicum*) en lugar de *B. floridae*, a saber [(*B. belcheri*+ *B. japonicum*)+ (*B. lanceolatum*+ *B. floridae*)+ *Asymmetron*]. Esta incoherencia puede deberse a la escasa información sobre el lugar de la secuencia y al uso de marcadores moleculares inadecuados (Zhong et al., 2009). Kon et al. (2007) encontraron la nueva especie de anfioxo *Asymmetron inferum*, cerca de cadáveres de ballenas en el fondo marino a 229 m de profundidad (cabo Nommisaki, suroeste de la isla de Kyushu, Japón) y utilizaron el ADNmt completo de ocho especies de anfioxo (incluidas otras tres especies como grupos externos) para reconstruir sus árboles bayesianos (BI), ML y de máxima parsimonia (MP). La topología fue coherente con la construida a través de 12S rRNA, en lugar de la basada en 13 PCGs por Zhong et al. (2009). Por lo tanto, la relación filogenética entre estas cuatro especies de *Branchiostoma* sigue sin estar clara y esta divergencia causada por varios genes de ADNmt todavía requiere aclaración.

La datación molecular basada en el árbol filogenético se ha investigado ampliamente dentro de las especies de anfioxus. Kon et al. (2007) estimaron el tiempo de divergencia de ocho especies de anfioxus utilizando métodos de datación molecular, mostrando que los géneros *Asymmetron* y *Branchiostoma* se separaron de su ancestro común (*Asymmetron* - *Branchiostoma*) alrededor de 240 Mya, y *B. belcheri* - *Branchiostoma* alrededor de 240 Mya.

B. floridae alrededor de 130 Mya. Sin embargo, más recientemente, Yue et al. (2014) construyeron un árbol ML de 15 especies basado en 427 ortólogos utilizando datos transcriptómicos y genómicos e incluyendo dos especies de amphioxus (*B. floridae* y *A. lucayanum*). Sugirieron un tiempo de divergencia de *Asymmetron* - *Branchiostoma* de unos 120 Mya. Huang et al. (2014) realizaron la secuenciación del genoma de *B. belcheri* (el segundo genoma de anfioxo después del de *B. floridae*) y estimaron los tiempos de divergencia de 15 especies basándose en 513 ortólogos de los conjuntos genómicos; sus datos cubrían las dos especies de anfioxo, *B. floridae* y *B. belcheri*, que predijeron 130 Mya como tiempo de divergencia entre ellas. Hasta ahora, se han investigado árboles filogenéticos que contenían múltiples especies de amphioxus basados en el ADNmt; sin embargo, los construidos recientemente mediante datos transcriptómicos y genómicos a gran escala contenían sólo dos especies de amphioxus (no más de dos especies). Por lo tanto, es necesario reconstruir un árbol filogenético que contenga múltiples especies de amphioxus y calcular el marco temporal de divergencia a nivel ómico para desvelar mejor la historia de la divergencia y paliar las contradicciones entre las investigaciones anteriores.

Además, investigaciones previas realizaron una comparación directa de la diversidad de dominios entre amphioxus y otros vertebrados modelo, identificando dominios proteicos específicos de amphioxus de *B. belcheri* y *B. floridae* (Huang et al., 2014). Sin embargo, los dominios específicos de linaje entre múltiples especies de anfioxus y su función biológica siguen siendo en gran parte desconocidos, lo que obstruye nuestra comprensión de las características de la especiación de los anfioxus.

Aquí secuenciamos el transcriptoma de *B. japonicum* adulto y ensamblamos, respectivamente, su transcriptoma y el de *A. lucayanum*. En combinación con el conjunto de genes del genoma completo de las otras 16 especies y el transcriptoma de *B. lanceolatum*, realizamos análisis comparativos a gran escala para reconstruir el árbol filogenético y estimar los tiempos de divergencia entre los cefalocordados. Además, se identificaron dominios específicos de cada especie y ancestrales

entre los anfioxus y se realizaron análisis de enriquecimiento funcional para cada conjunto de dominios diana.

MATERIALES Y MÉTODOS

Declaración ética

Este estudio se llevó a cabo de acuerdo con las recomendaciones de la Guía para el Cuidado y Uso de Animales Invertebrados de Laboratorio. El protocolo fue aprobado por el Comité Ético de Investigación de la Universidad de Nanjing (NJU).

Materiales y secuenciación del ARN

Se recogieron ejemplares adultos de *B. japonicum* en Qingdao, provincia de Shandong (China), y se mantuvieron en el laboratorio durante 5 días para facilitar el vaciado del tubo digestivo. Se aisló el ARN total de varios machos y hembras utilizando el reactivo TRIzol (Invitrogen, Estados Unidos), y el ADN genómico se eliminó mediante RNase-Free DNase Set (Qiagen, Alemania) siguiendo las instrucciones del fabricante. La calidad del ARN se determinó con el Bioanalyser 2100 (Agilent, Estados Unidos), y la concentración se midió con el NanoDrop 1000 (Thermo Scientific, Estados Unidos). Las bibliotecas de ADNc estándar se construyeron utilizando los protocolos estándar de preparación de bibliotecas de Illumina y el kit TruSeq (Illumina, Estados Unidos). Tras el tratamiento de homogeneización de las bibliotecas de secuenciación, se llevó a cabo la secuenciación del ARN (ARN-seq) en un Illumina HiSeqTM 2000 con lecturas de 100 pb de extremo pareado. La ARN-seq se realizó en BGI (Shenzhen, China).

Ensamblaje de lecturas, corrección de modelos genéticos y anotación funcional

En primer lugar, utilizamos FastQC (Brown et al., 2017) para controlar la calidad de las lecturas de secuenciación en bruto mediante la comprobación de secuencias sobre-representadas y potencialmente contaminantes siguiendo criterios estrictos: (1) se descartaron las lecturas con adaptadores;

(2) se descartaron las lecturas con bases desconocidas > 10%; (3) se descartaron las lecturas con una longitud < 20 pb. Se utilizó el programa Sickle para eliminar o recortar las lecturas de baja calidad [porcentaje de bases de baja calidad (bases con puntuación de calidad de secuenciación ≤ 5) > 50%] (Cabili et al., 2011). Se obtuvieron lecturas pareadas de calidad (lecturas limpias) y se utilizaron para ensamblar el transcriptoma utilizando Trinity (Grabherr et al., 2011). Además, obtuvimos lecturas limpias de otro anfioxo adulto (*A. lucayanum*) (Accession: SRX437621) de la base de datos Sequence Read Archive (SRA) del NCBI, y ensamblamos su transcriptoma siguiendo los procedimientos anteriores. Trinity puede detectar isoformas potenciales de splicing alternativo y etiquetarlas con el mismo prefijo. En el caso de isoformas múltiples, se seleccionó el unigén más largo de cada grupo de isoformas como representación única de ese grupo. También descargamos el transcriptoma de *B. lanceolatum* de la base de datos NCBI Transcriptome Shotgun Assembly (TSA) (Números de acceso: JT846176-JT905674) (Silvan et al., 2012). Las posibles secuencias codificantes (CDS) de tres especies de amphioxus se predijeron mediante la búsqueda Blastx (parámetro por defecto) (Mount, 2007) en la base de datos de proteínas no redundantes (NR) del NCBI. El alineamiento óptimo se extrajo como plantilla para la determinación de las

TABLA 1| Estadísticas resumidas de secuenciación, ensamblaje y anotación de transcriptomas.

	<i>B. japonicum</i>	<i>A. lucayanum</i>
Lecturas en bruto	~52 millones de euros	~146 millones de euros
Lecturas limpias	~49 millones	~135 millones de euros
Q20 porcentaje	97.09%	/
N50 (pb)	1,753	1,797
Número de unigenes finales	92,003	112,753
Número de CDS	50,316	42,148
NR	46,540	/
Swiss-Prot	31,941	/
IR A	30,319	/

CDS de los unigenes. A continuación, se utilizó el software ESTScan (valor $E = 10^{-5}$) (Iseli et al., 1999) para predecir las CDS de los unigenes que no produjeron ningún acierto.

Las secuencias de proteínas, ARNm y CDS de 16 especies muestreadas con genoma secuenciado se recuperaron de bases de datos en línea (véase la **Tabla suplementaria S1** para más detalles). Entre estas secuencias genómicas, las secuencias CDS y proteicas de 15 especies (excepto *A. mississippiensis*) no pudieron coincidir completamente, por lo que se corrigieron. En primer lugar, se eliminaron las secuencias CDS inferiores a 120 pb; a continuación, las secuencias proteicas se alinearon con el ARNm utilizando el software Exonerate (Guy y Birney, 2005). Basándose en las alineaciones aminoácido-ácido nucleico, se corrigieron las secuencias CDS y proteicas incompatibles, descartando las secuencias no corregidas. Para las secuencias proteicas traducidas a partir de las CDS corregidas, utilizamos la herramienta Blastp para buscarlas en las bases de datos Nr, Swiss-Prot y de ontología génica (GO) (**Tabla 1**), respectivamente, realizando la anotación funcional de todos los genes o unigenes.

Construcción de alineamientos ortólogos

Los métodos HaMSTR (Ebersberger et al., 2009) y de mejores coincidencias recíprocas (BRH) (Moreno-Hagelsieb y Latimer, 2008) se utilizan generalmente para la construcción de genes ortólogos. Los resultados obtenidos a partir de los métodos HaMSTR y BRH se intersectaron como un conjunto ortólogo final de 19 especies muestreadas, para evitar el sesgo causado por el uso de un solo método. Para el enfoque HaMSTR, de acuerdo con la descripción previa en los análisis transcriptómicos (Misof et al., 2014; Li et al., 2015), se construyó un conjunto de core-orthologs a partir de nueve genomas de vertebrados incluyendo el humano (*Homo sapiens*), ratón (*Mus musculus*), pollo (*Gallus gallus*), pez cebra (*Danio rerio*), celacanto (*Latimeria chalumnae*), fugu (*Takifugu rubripes*), tiburón elefante (*Callorhynchus milii*), lamprea (*Petromyzon marinus*) y rana de uñas occidental (*Xenopus tropicalis*). Las 6.793 proteínas core-ortólogas se obtuvieron de la base de datos OrthoDB v9.1 (Kriventseva et al., 2015) y se introdujeron en la herramienta de alineamiento múltiple Muscle (v3.8) (Edgar, 2004) utilizando conjuntos predeterminados. A continuación, se construyó un modelo de Markov oculto "primer taxa" a partir del alineamiento múltiple de los core-orthologs utilizando la herramienta hmbuild del paquete HMMER3 (Finn et al., 2011). El "primer taxa" sirvió como entrada para producir la base de datos core-ortholog para el programa HaMSTR v.13.2.6 (Ebersberger et al., 2009) para buscar los correspondientes orthologs en las 10 especies restantes.

Si una especie contenía múltiples ortólogos correspondientes en co-ortólogos, sólo se retuvo el acierto óptimo. HaMSTR v.13.2.3 se ejecutó con parámetros estrictos (-sequence file, -est, -hmmset, -refspec, -representative y -ublast) de acuerdo con una descripción anterior (Li et al., 2015). Además, se empleó Proteinortho (v5.13) (valor $E = 10^{-10}$) (Lechner et al., 2011), un software compilado basado en el método BRH que se utiliza ampliamente para el análisis genómico comparativo a gran escala, para identificar co-ortólogos entre las 19 especies. Se utilizó el programa PhyloTreePruner para seleccionar el representante único de cada especie en los genes co-ortólogos basándose en el árbol genético (Kocot et al., 2013). De acuerdo con las etiquetas de los genes, los grupos de genes ortólogos obtenidos se incluyeron colectivamente en los resultados de los métodos HaMSTR y BRH. Todos los genes ortólogos, incluidas las secuencias de proteínas y CDS, se alinearon mediante la herramienta Prank (opción "-codon") (Löytynoja y Goldman, 2005) y, a continuación, se recortaron mediante Gblocks (Talavera y Castresana, 2007) con el parámetro "-t = c" para eliminar las regiones mal alineadas. Se eliminaron las alineaciones recortadas que contenían menos de 60 pb/20 aminoácidos.

Análisis filogenético

Las alineaciones concatenadas utilizadas para la construcción del árbol filogenético se construyeron a partir de todos los ortólogos utilizando el paquete FasParser (Sun, 2017). El mejor modelo de sustitución de aminoácidos se identificó para el posterior análisis filogenético utilizando ProtTest (v3.4) (Abascal et al., 2005). El modelo LG I G++ + F se recomendó como el mejor modelo basado tanto en el criterio de información de Akaike (AIC) como en el criterio de información bayesiano (BIC). Se llevaron a cabo análisis filogenéticos para todos los conjuntos de datos de proteínas alineados utilizando los modelos sitio-homogéneo (ML y Bayes) y

-modelos heterogéneos (PhyloBayes). Se utilizó RAxML 7.0.4 (Guindon et al., 2010) para construir árboles ML, y se utilizaron 1000 bootstraps (BS) para estimar la fiabilidad de los nodos. El árbol filogenético bayesiano se reconstruyó utilizando MrBayes 3.2.2 (Ronquist et al., 2012). En el análisis MrBayes, el modelo LG no estaba disponible, por lo que se utilizó el modelo JTT I G++ + F como mejor opción secundaria (Yue et al., 2014). Se realizaron simultáneamente dos ejecuciones independientes con cuatro cadenas (tres calentadas y una fría) durante 1.000.000 de generaciones, muestreando cada 100 generaciones. Cuando el valor del tamaño estimado de la muestra (ESS) superó 100 y el factor de reducción de escala potencial (PSRF) se acercó a 1,0, se consideró que se había alcanzado la estacionariedad, tal y como recomienda el software MrBayes (Ronquist et al., 2012). A continuación, se descartó el primer 25% de las muestras como burn-in, y se calcularon las longitudes de rama y las probabilidades posteriores (PP) en un árbol de consenso. Los análisis bayesianos con un modelo de sitios heterogéneos se implementaron utilizando PhyloBayes 4.1b (Lartillot et al., 2013). Tras la eliminación de los sitios constantes del alineamiento, se ejecutaron dos cadenas independientes, partiendo de un árbol aleatorio, bajo el modelo CAT-GTR.

A continuación, también utilizamos ASTRAL II, un algoritmo estadísticamente consistente para estimar la topología del árbol de especies bajo el modelo de coalescencia multispecie (Mirarab et al., 2016). Los valores de soporte se obtuvieron calculando la probabilidad posterior local, una ventaja de ASTRAL II que presentó una alta precisión en comparación con el bootstrapping multilocus en un amplio conjunto de

conjuntos de datos simulados y biológicos (Sayyari y Mirarab, 2016). Realizamos el análisis sin el conjunto "mapa de especies", lo que significa que no es necesario asignar múltiples individuos de la misma especie a un taxón. Árbol de especies obtenido a partir de los árboles de genes RAxML de cada uno de los 3070 loci ortólogos, resumidos con ASTRAL II. Como los análisis de árboles de especies no requieren outgroups (Heled y Drummond, 2009), las 19 especies se incluyeron como ingroups en los análisis ASTRAL II.

Estimación del tiempo de divergencia

Basándose en el árbol ML reconstruido anteriormente, se realizó una estimación bayesiana del tiempo de divergencia de las 19 especies muestreadas para construir un marco temporal de la evolución de los cefalocordados. Los nueve nodos de divergencia se señalaron como puntos de calibración de acuerdo con registros fósiles previamente resumidos (Benton y Donoghue, 2007; Yue et al., 2014). Se utilizó un punto de calibración (520,00-? Mya) de nuestras primeras investigaciones en Olfactores (n) (Chen et al., 1999). En el análisis de datación molecular, un límite superior seguro para la edad raíz es esencial en el Árbol de Monte Carlo con cadena de Markov (MCMCTree). Dado que no existe ningún registro fósil fiable que pudiera utilizarse aquí, se utilizó el enfoque de calibración secundaria para obtener la edad raíz (q, Deuterostomia). Actualmente, el tiempo de divergencia en el nodo de Deuterostomia se estimó basándose en el árbol con el nodo raíz de Bilateria (Deuterostomia+ Mollusca) a través del reloj molecular. Sin embargo, se han notificado varios valores para las restricciones máximas del tiempo de divergencia de los bilaterios coronados, incluyendo 581,5 Mya (Benton y Donoghue, 2007), y 640-730 Mya (Peterson et al., 2008). Yue et al. (2014) exploraron el tiempo de divergencia de los bilaterios de la corona mediante el análisis MCMCTree, incluyendo calibraciones de fósiles bien descritas, bajo límites inferiores de raíces fijas (531,50) y una variedad de límites superiores (600, 700, 800 y 900 Mya). A continuación, los autores propusieron que 600 y 700 Mya eran fiables y calcularon su respectivo tiempo de divergencia de Deuterostomia basándose en ambas restricciones de calibración (dando como resultado 532,67-598,27 y 585,50-698,50 Mya) (Yue et al., 2014). En consecuencia, empleamos ambos resultados como diferentes fechas de calibración secundarias para el nodo Deuterostomia (q) en las investigaciones actuales. Todas las restricciones de calibración de los nodos de divergencia se presentan en la **Tabla 2**.

El análisis de datación molecular se realizó utilizando el subprograma MCMCTree del paquete PAML (v4.8) con la matriz de genes concatenados utilizada en el análisis filogenético (Dos y Yang, 2011). Como topología fija del árbol guía, primero obtuvimos las estimaciones ML de las longitudes de rama, el vector gradiente (G) y la matriz hessiana (H) utilizando el codeml del paquete PAML (v4.8) (Dos y Yang, 2011), después, estos valores estimados se introdujeron en los archivos de control de MCMCTree junto con el modelo de sustitución LG. Establecimos una unidad temporal de 1 millón de años y utilizamos el mismo previo gamma G(1, 100) para especificar los parámetros de tasa de sustitución y deriva de la tasa. Se utilizó un modelo de reloj relajado autocorrelacionado y lognormal (Rannala y Yang, 2007) para estimar la distribución posterior del tiempo de divergencia, dadas estas priorizaciones. El burn-in para la cadena MCMC se ejecutó 10.000.000 de generaciones, con parámetros de muestreo cada 1.000 generaciones. realizaron dos ejecuciones independientes para garantizar la convergencia de las cadenas MCMC. El programa resumió la media y los intervalos de confianza (IC) del 95%.

TABLA 2 | Estimaciones Bayesianas MCMC y R8S para el tiempo de divergencia de cada nodo interno mostrado en la Figura 2, asumiendo que la divergencia de Deuterostomia (restricciones de calibración de la raíz) ocurrió en [532.67, 598.27] Mya.

Índice de nodo	Nombre del nodo	Restricciones de calibración	MCMCTREE		R8S	
		[mín, máx] (Mya)	Media (Mya)	IC 95% (Mya)	Media (Mya)	IC 95% (Mya)
a	Eutheria	[61.50, 100.50]	76.49	[60.91-99.43]	71.65	[56.72-97.68]
b	Amniote	[312.30, 330.40]	318.63	[312.00-329.34]	312.67	[301.47-326.36]
c	Ankylopoda	[259.70, 299.80]	268.28	[258.33-291.27]	257.36	[253.21-284.63]
d	Archosauriformes	[235.00, 250.40]	244.31	[235.42-250.63]	246.28	[237.34-249.69]
e	Tetrápodos	[330.40, 350.10]	343.13	[331.80-350.42]	344.07	[331.78-351.51]
f	Sarcopterygii	-	382.15	[354.82-407.11]	367.22	[333.78-398.61]
g	Osteichthyes	[416.00, 421.75]	418.68	[416.02-421.73]	424.82	[418.13-431.44]
h	Acanthopterygii	[149.85, 165.20]	157.36	[149.81-165.20]	152.92	[146.65-166.62]
i	Gnathostomata	[412.75, 462.50]	443.11	[420.23-461.68]	448.65	[413.27-477.76]
j	Vertebrata	[460.60, -]	494.47	[460.94-532.75]	471.19	[452.53-498.87]
k	Olfactores	[520.00, -]	548.41	[513.66-579.60]	525.77	[504.49-546.35]
l	Cordados	-	577.52	[544.32-597.51]	546.06	[502.15-588.26]
m	-	-	61.11	[9.93-128.65]	64.23	[12.58-123.75]
n	Branchiostomidae	-	86.83	[39.98-149.47]	91.11	[32.84-162.85]
o	-	-	72.42	[30.50-140.35]	78.64	[34.22-143.95]
p	Cephalochordata	-	104.37	[21.25-191.35]	112.09	[28.80-193.76]
q	Deuterostomia	[532.67, 598.27]	587.14	[557.15-601.25]	569.53	[543.25-589.17]
r	Ambulacraria	-	569.65	[461.45-629.35]	547.27	[414.91-601.75]

El enfoque de probabilidad penalizada (PL) se utilizó en el programa r8s V1.8 (Sanderson, 2003), en combinación con el algoritmo truncado de Newton (Sanderson, 2002). El método de validación cruzada en r8s se utilizó para determinar el nivel óptimo de suavizado de la tasa de los análisis PL con parámetros de suavizado que variaban de 1 a 1.000 según un procedimiento descrito previamente (Mulcahy et al., 2012). El parámetro de suavizado óptimo ($S = 100$) se determinó primero utilizando la misma topología fija (ML) utilizada como entrada para MCMCTree, ya que existen pruebas de que las longitudes de rama se estiman con mayor precisión mediante ML que mediante métodos BI (Schwartz y Mueller, 2010). Los IC para las estimaciones de edad de PL se obtuvieron replicando los análisis de PL de 1.000 árboles. El tiempo de divergencia medio y los IC del 95% se resumieron utilizando el kit bootstrap r8s.

Análisis de la evolución molecular

Para identificar los dominios contenidos en las secuencias de proteínas de todos los anfioxus, se buscaron los conjuntos de secuencias de aminoácidos de cinco especies de anfioxus en la base de datos Pfam-A 31.0 utilizando el programa local HMMER 3.1¹. Este es el mejor programa para la identificación de dominios y se ha utilizado a menudo junto con una base de datos de perfiles, como Pfam (Rekapalli et al., 2009). El conjunto de identificación de dominios de cada anfioxo se introdujo en un diagrama de Venn² para obtener tipos de dominios específicos y compartidos entre las cinco especies de anfioxos investigadas. Cabe señalar que sólo los conjuntos de proteínas de *B. belcheri* y *B. floridae* se generaron a partir del genoma completo secuenciado, mientras que los de las otras tres especies se obtuvieron a partir de transcriptomas adultos ensamblados *de novo* que contenían defectos de ensamblaje y secuencias de proteínas faltantes.

¹ <http://www.hmmerr.org/>

² <https://cran.r-project.org/>

Sin embargo, un tipo de dominio podía estar contenido en una variedad de secuencias de proteínas y el tamaño del dominio era corto (generalmente en el rango de 40-50 aminoácidos), lo que evitaba eficazmente el sesgo causado por conjuntos incompletos de proteínas de especies y errores de ensamblaje *de novo*. Las secuencias de dominio se anotaron en bases de datos GO utilizando la herramienta Blastp con el valor E por defecto. El análisis de enriquecimiento GO para dominios específicos y compartidos entre todos los anfioxus se realizó mediante el test exacto de Fisher, que se implementó en el pipeline Blast2GO (Conesa et al., 2005). Se utilizó el método Benjamini-Hochberg (BH) para la corrección de la tasa de falsos descubrimientos (FDR) para la prueba de Fisher (valores umbral de FDR = 0,05). La lista de términos GO significativos se filtró adicionalmente utilizando GO trimming (v2.0) (Jantzen et al., 2011) para descartar términos redundantes. En la **Figura 1A** se presenta el pipeline del software primario utilizado en este estudio.

RESULTADOS

Secuenciación, ensamblaje y anotación

La secuenciación Illumina para *B. japonicum* generó ~ 52 millones de lecturas en bruto. Tras el control de calidad de los datos originales, se obtuvieron ~ 49 millones de lecturas limpias, y el porcentaje Q20 alcanzó el 97,09%, lo que indica una alta calidad de la secuenciación del transcriptoma (**Tabla 1**). Los datos limpios se han enviado a la base de datos NCBI SRA: Número de acceso SRX448311. Mediante el ensamblaje inicial de Trinity, obtuvimos 161.542 contigs con un valor N50 de 816 pb y una longitud media de 375 pb. Tras el agrupamiento por solapamiento de las secuencias contig, obtuvimos finalmente 92.003 unigenes con un valor N50 de 1.753 pb y una longitud media de 785 pb. Los resultados del ensamblaje de *A. lucayanum* fueron casi consistentes con informes previos que secuenciaron originalmente sus transcriptomas (Yue et al., 2014). Además, alineamos

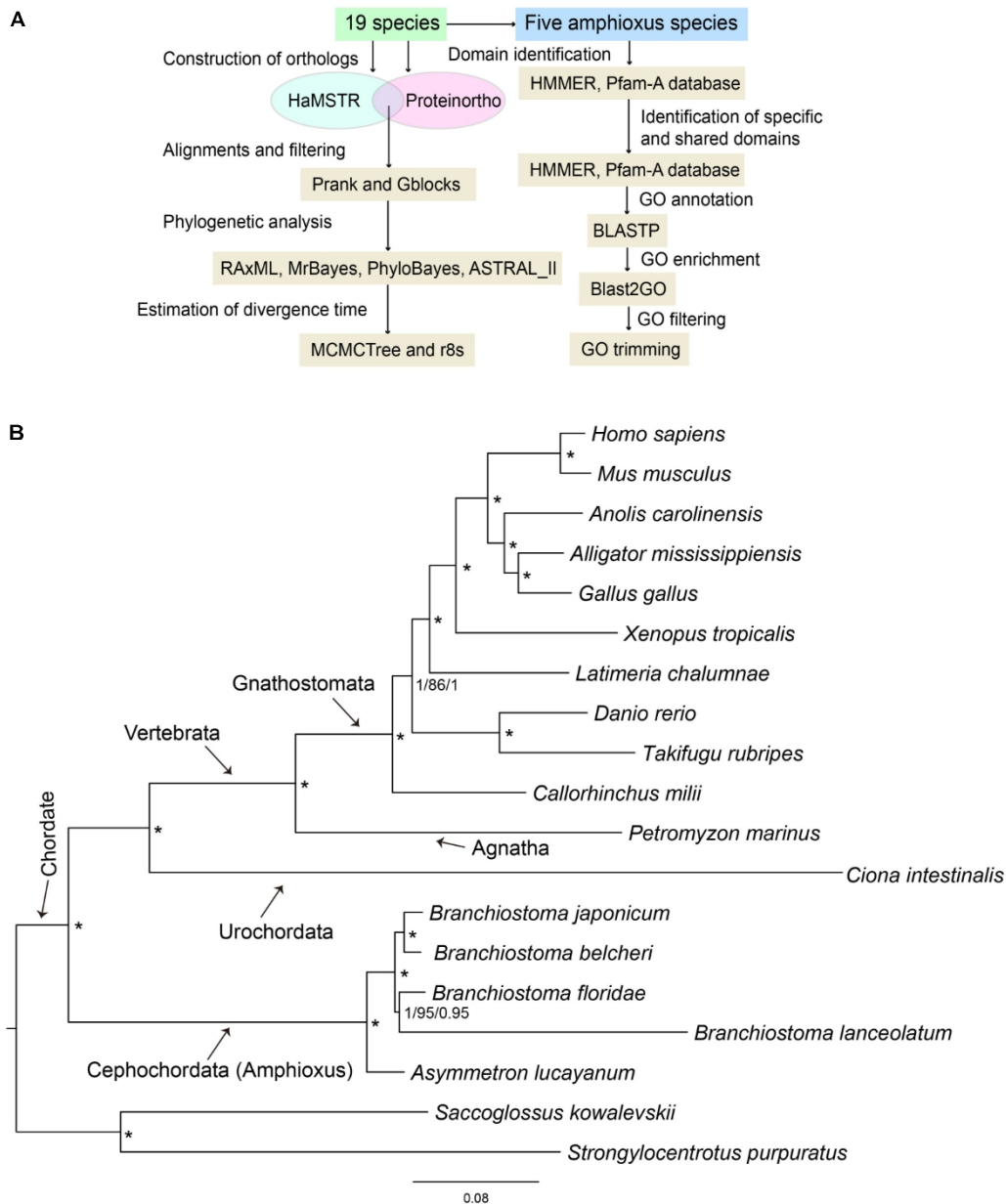


FIGURA 1| (A) Esquema del software primario utilizado en este estudio. (B) Árbol de máxima verosimilitud (ML) de 19 vías inferido a partir de los 397 grupos de genes ortólogos concatenados. La longitud de la rama es proporcional a la tasa de sustitución de aminoácidos esperada y la barra de escala representa 0,08 sustituciones de aminoácidos esperadas por sitio. Los números de los nodos internos representan las probabilidades posteriores bayesianas (PP1, izquierda) de MrBayes, los valores bootstrap (BS, centro) de RAxML y las probabilidades posteriores bayesianas (PP2, derecha) de PhyloBayes. Un asterisco indica PP1= 1.0, BS= 100, y PP2= 1.0. Los árboles bayesianos y de PhyloBayes se muestran en las Figuras suplementarias S1, S2, respectivamente.

con secuencias unigénicas de *B. japonicum*, el 86,15% de las lecturas limpias pudieron emparejarse completamente, lo que indica que la calidad del ensamblaje cumple los requisitos para el análisis posterior. A partir del conjunto de unigenes obtenido, se obtuvieron 46.145 y 4.171 secuencias CDS mediante búsqueda de alineación homóloga y detección ESTScan, respectivamente. Se obtuvieron 46.540, 31.941 y 30.319 unigenes de

B. japonicum se anotaron en NR, Swiss-Prot, base de datos GO (Tabla 1). Para las secuencias genómicas corregidas, sus secuencias CDS son totalmente idénticas a sus proteínas correspondientes, garantizando

fiabilidad de los análisis posteriores. Las estadísticas resumidas de la corrección de genes se muestran en la Tabla Suplementaria S1.

Tasa de evolución y filogenia molecular de Amphioxus

Para aclarar aún más las relaciones filogenéticas entre los anfioxos, realizamos un análisis filogenético con cinco especies de cefalocordados (*B. belcheri*, *B. japonicum*,

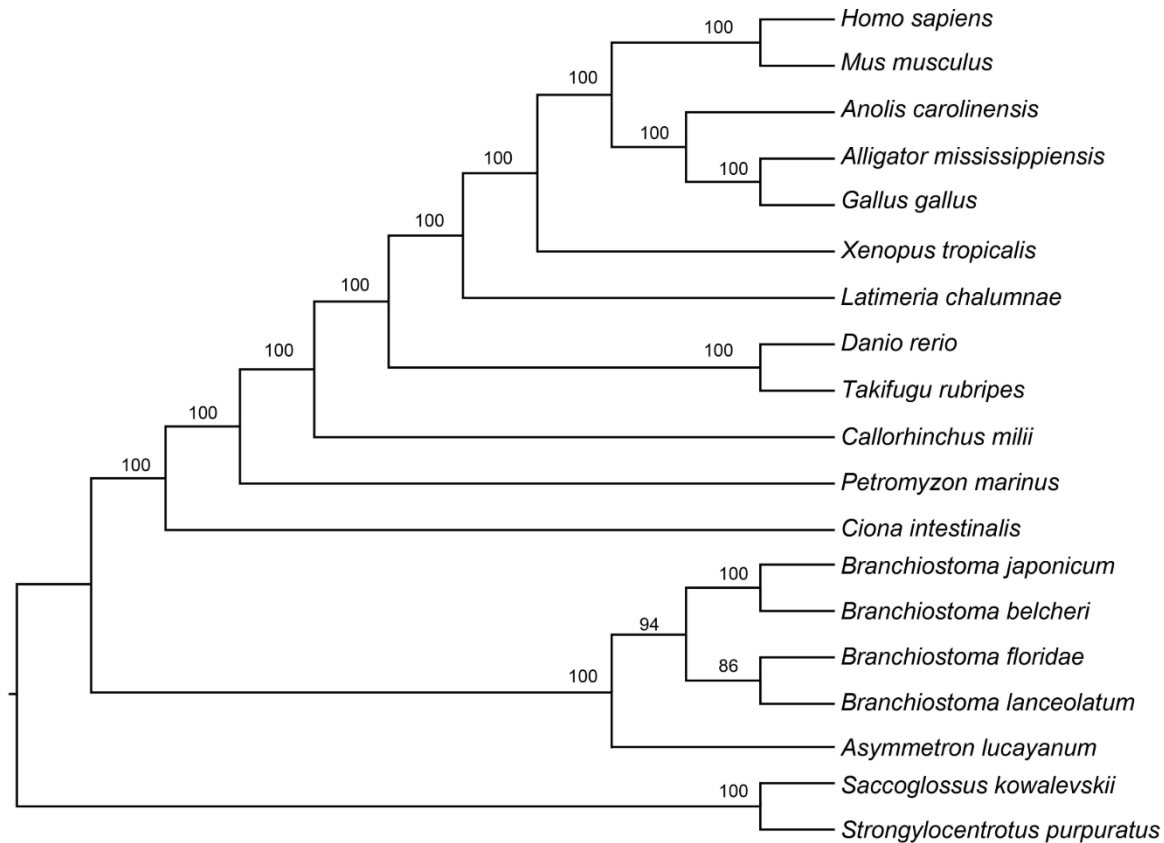


FIGURA 2| Resultados de los análisis del árbol de especies a partir del conjunto de datos de 19 especies. El soporte de las ramas se estimó calculando la probabilidad posterior local.

B. lanceolatum, *B. floridae* y *A. lucayanum*), 11 especies de vertebrados (humano, ratón, caimán americano, lagarto verde, pollo, rana de uñas occidental, celacanto, fugu, pez cebra, tiburón elefante y lamprea), un urocordado (ascidia) y dos grupos externos (gusano bellota y erizo de mar). Para construir el árbol filogenético se utilizó un alineamiento múltiple concatenado de 19 proteínas, basado en 397 grupos de genes ortólogos intersectados por los métodos HaMSTR y BRH (442 para HaMSTR y 415 para BRH). Además, la longitud de rama (la tasa de sustitución de aminoácidos esperada) se calculó para cada especie basándose en ML (Figura 1B), BI (Figura S1Suplementaria Figura Suplementaria S2) y PhyloBayes (). Estos tres métodos de inferencia obtienen topologías de árbol congruentes con longitudes de rama casi idénticas (Figura 1B); los valores de soporte nodal son generalmente altos (PP= 0,95-1,0, BS= 86-100); además, estos valores son más fuertes en el árbol BI que en el árbol ML, como se ha informado anteriormente (Yuan et al., 2015; Yuan et al., 2016). El árbol de especies estimado por ASTRAL-II es muy similar al árbol estimado en el análisis concatenado, con la mayoría de los nodos mostrando el 100% de la probabilidad posterior local (Figura 2). Independientemente del método analizado (ML, BI o PhyloBayes) utilizado, las longitudes de rama de todos los anfioxos excepto *B. lanceolatum* fueron consistentemente más cortas que las de los vertebrados, incluso más cortas que la del tiburón elefante (*Callorhynchus milii*) como vertebrado de evolución más lenta (Venkatesh et al., 2014). El mar

squirt mostró la mayor longitud de rama en el árbol filogenético. Curiosamente, la longitud de rama de *B. lanceolatum* es mayor que la de otros anfioxos y vertebrados investigados en este estudio. Sobre la base de esta diferencia en la longitud de la rama, la tasa de sustitución de aminoácidos de *B. lanceolatum* se evaluó para ser ~ 1,3 veces mayor que la de fugu (*Takifugu rubripes*) con la tasa evolutiva más rápida entre los vertebrados. Realizamos la prueba de tasa relativa de Tajima (Tajima, 1993) basándonos directamente en la alineación concatenada de secuencias de proteínas de las pruebas por pares entre *B. lanceolatum* y fugu, que demostró un resultado significativo (valor *P* inferior a 0,05). Además, los árboles filogenéticos apoyan sistemáticamente una filogenia de [(*B. belcheri*+ *B. japonicum*)+ (*B. lanceolatum*+ *B. floridae*+ *A. lucayanum*)], independientemente del método analítico utilizado.

Estimaciones del tiempo de divergencia entre las especies de Amphioxus

Basándonos en el árbol ML, realizamos un análisis MCMCTree para los tiempos de divergencia entre las 19 especies analizadas (Figura 3). Se utilizaron once restricciones temporales en todo el árbol filogenético (véase la sección "Materiales y métodos"). Los tiempos de divergencia de Deuterostomia (edad raíz) se emplearon como 532,67-598,27 y 585,50-698,50 Mya, respectivamente, según investigaciones anteriores (Yue et al., 2014). Los resultados de todas las estimaciones de edad con un IC del 95

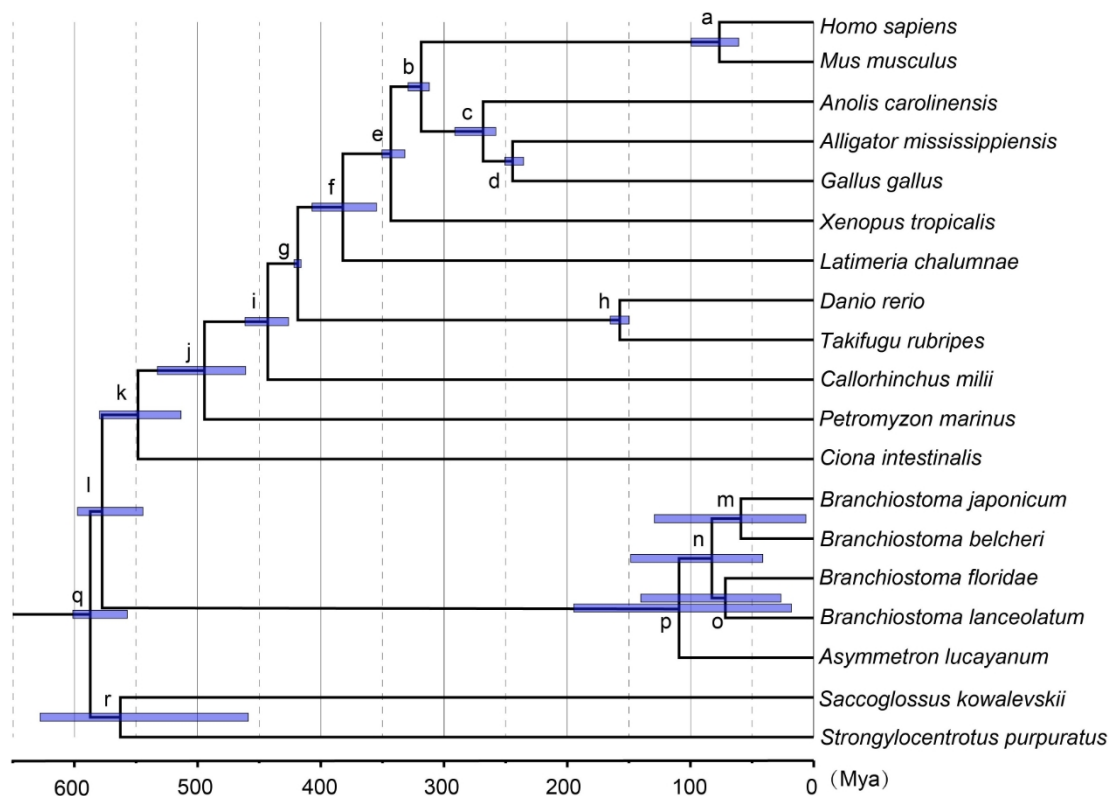


FIGURA 3| Estimaciones del marco temporal de divergencia de la evolución de los cefalocordados utilizando MCMCTree. Las letras minúsculas sobre la barra lila representan etiquetas de nodos internos (nodos a-r). Todas las limitaciones temporales correspondientes, los tiempos de divergencia medios y los intervalos de confianza (IC) del 95% obtenidos por las estimaciones bayesianas se tabulan en la Tabla 2. Cada barra lila representa el IC del 95% de la estimación correspondiente.

(intervalo de confianza) y los valores medios se muestran en las Tablas 2, 3. Independientemente de la edad raíz utilizada, se estimó que el ancestro común más reciente de las especies de anfioxos vivientes (el tiempo de divergencia entre *Asymmetron* y *Branchiostoma*) era de ~104 Mya; las primeras escisiones dentro de cada clado principal se produjeron en ~87 Mya; *B. lanceolatum* divergió de *B. floridae* en ~72 Mya, y el tiempo de divergencia entre *B. belcheri* y *B. japonicum* fue ~61 Mya. Además, para garantizar la solidez de los resultados de la estimación en el marco temporal de divergencia, se empleó el análisis R8S utilizando el mismo conjunto de restricciones temporales (con una edad raíz de 532,67- 598,27 Mya) que se utilizaron para la estimación MCMCTree. Todos los resultados de divergencia con intervalos de IC del 95% y valores medios se muestran en la Tabla 2. A pesar de los tiempos de divergencia obviamente diferentes en varios nodos, por ejemplo, los nodos f, k, q y r, los análisis r8s apoyaron en general los resultados del MCMCTree, particularmente dentro de los cefalocordados. En consecuencia, los tiempos de divergencia calculados con el MCMCTree eran fiables y se utilizaron directamente para la discusión posterior.

Identificación y análisis funcional de dominios proteicos compartidos y específicos de cada especie de Amphioxus

Los dominios son unidades evolutivas funcionales independientes que pueden plegarse independientemente (Vogel et al., 2004); la mayoría de ellos

tienen una historia original antigua (Itoh et al., 2007). Para investigar la función de los dominios antiguos y específicos de cada especie entre los anfioxos, en primer lugar recuperamos 10.106, 10.708, 8.912, 8.313 y 9.821 dominios de *A. lucayanum*, *B. belcheri*, *B. floridae*, *B. lanceolatum*, y *B. japonicum*, respectivamente, mediante la búsqueda de dominios homólogos en la base de datos Pfam-A. Entre estos dominios, 4.814 (grupo 1, G1) fueron compartidos entre los géneros *Asymmetron* y *Branchiostoma* (Figura Suplementaria S3) y fueron retenidos y transmitidos desde su ancestro común.

Además, el análisis de enriquecimiento de la ontología génica (GO) para G1 descubrió que 162 términos GO enriquecidos pertenecientes a la subcategoría "proceso biológico" están relacionados principalmente con el desarrollo, el proceso y la función celular, y los procesos metabólicos (Tabla suplementaria S2). Además, detectamos 492 (grupo 2, G2), 361 (grupo 3, G3), 427 (grupo 4, G4), 255 (grupo 5, G5) y 366 (grupo 6, G6) dominios específicos de especie en *B. belcheri*, *B. floridae*, *B. lanceolatum*, *B. japonicum* y *B. lanceolatum*.

A. lucayanum, respectivamente (Tablas Suplementarias S3-S7). Para examinar si estos dominios específicos de especie tenían funcionalidad biológica, también investigamos el enriquecimiento de términos GO para aquellos en cada uno de G2-6 (Tabla Suplementaria S8). Entre los 21 términos GO enriquecidos en G2 (*B. belcheri*), destacaron en particular los asociados a la respuesta inmune, inflamatoria, de estímulo, apoptosis y fagocitosis. En el caso de G3 (*B. floridae*), se enriquecieron 42 términos GO, destacando los asociados al desarrollo,

TABLA 3| Estimaciones Bayesianas MCMC para los tiempos de divergencia de cada nodo interno mostrado en la Figura 2, asumiendo que la divergencia de Deuterostomia (restricciones de calibración de la raíz) ocurrió en [585.50-698.50] Mya.

Nodo índice	Nombre del nodo	Calibración restricciones	MCMCTREE	
		[Mín, Máx] (Mya)	Media (Mya)	IC 95% (Mya)
a	Eutheria	[61.50, 100.50]	73.34	[60.85-98.57]
b	Amniote	[312.30, 330.40]	318.57	[311.92-329.16]
c	Ankylopoda	[259.70, 299.80]	267.54	[257.93-290.14]
d	Archosauriformes	[235.00, 250.40]	244.41	[235.51-250.75]
e	Tetrápodos	[330.40, 350.10]	343.41	[332.18-350.56]
f	Sarcopterygii	-	382.33	[356.45-406.58]
g	Osteichthyes	[416.00, 421.75]	418.74	[416.16-421.72]
h	Acanthopterygii	[149.85, 165.20]	157.37	[149.85-165.11]
i	Gnathostomata	[412.75, 462.50]	447.07	[428.73-462.66]
j	Vertebrata	[460.60, -]	530.01	[477.45-592.12]
k	Olfactores	[520.00, -]	616.81	[554.53-667.15]
l	Cordados	-	662.19	[603.55-697.00]
m	-	-	60.93	[10.21-136.41]
n	Branchiostomidae	-	84.92	[41.02-162.19]
o	-	-	71.71	[31.93-149.68]
p	Cephalochordata	-	100.11	[23.47-211.63]
q	Deuterostomia	[585.50-698.50]	677.64	[628.58-709.37]
r	Ambulacraria	-	601.93	[478.94-712.88]

inmunidad y apoptosis. A continuación, 33 términos GO enriquecidos de G4 (*B. lanceolatum*) estaban relacionados principalmente con el almacenamiento y la regulación de lípidos, la apoptosis, la diferenciación y el movimiento de células inmunitarias, y la respuesta a estímulos. En G5 (*B. japonicum*), 18 términos GO enriquecidos estaban relacionados principalmente con la oxidación y el metabolismo de las lipoproteínas, los movimientos antioxidantes, la apoptosis y la respuesta inmunitaria. Los 37 términos GO enriquecidos en G6 (*A. lucayanum*) estaban relacionados con el metabolismo orgánico y energético, el desarrollo muscular y el proceso del sistema inmunitario. En general, los términos GO relacionados con la inmunidad innata y la apoptosis fueron los más comúnmente enriquecidos por estos dominios específicos de la especie, seguidos por los términos GO relacionados con el metabolismo y la regulación de los lípidos, así como con el desarrollo de los tejidos, lo que sugiere un papel central de estos dominios relacionados en la especiación de los amphioxus.

DEBATE

Este estudio analizó datos ómicos a gran escala de múltiples especies de amphioxus (> 2 especies) para estimar su filogenia molecular. Nuestros análisis filogenómicos, basados en el conjunto de datos de supergenes con cuatro métodos analíticos, incluidos tres árboles de genes y un árbol de especies, obtuvieron una filogenia altamente respaldada de [(*B. belcheri*+ *B. japonicum*)+ (*B. lanceolatum*+ *B. floridae*)+ *A. lucayanum*]. Este resultado era coherente con análisis anteriores basados en el ARNr 12S (Zhong et al., 2009) y las secuencias completas de ADNmt (Kon et al., 2007); sin embargo, eran incongruentes con los resultados obtenidos con 13 PCG, que

indicaron una relación filogenética de (((*B. belcheri*+ *B. japonicum*)+ *B. lanceolatum*)+ *B. floridae*)+ *A. lucayanum*). Esta discrepancia es probablemente causada por diferentes presiones evolutivas y/o de selección, como se ha informado para otros animales (Havird y Santos, 2014; Yuan et al., 2018). Varios estudios

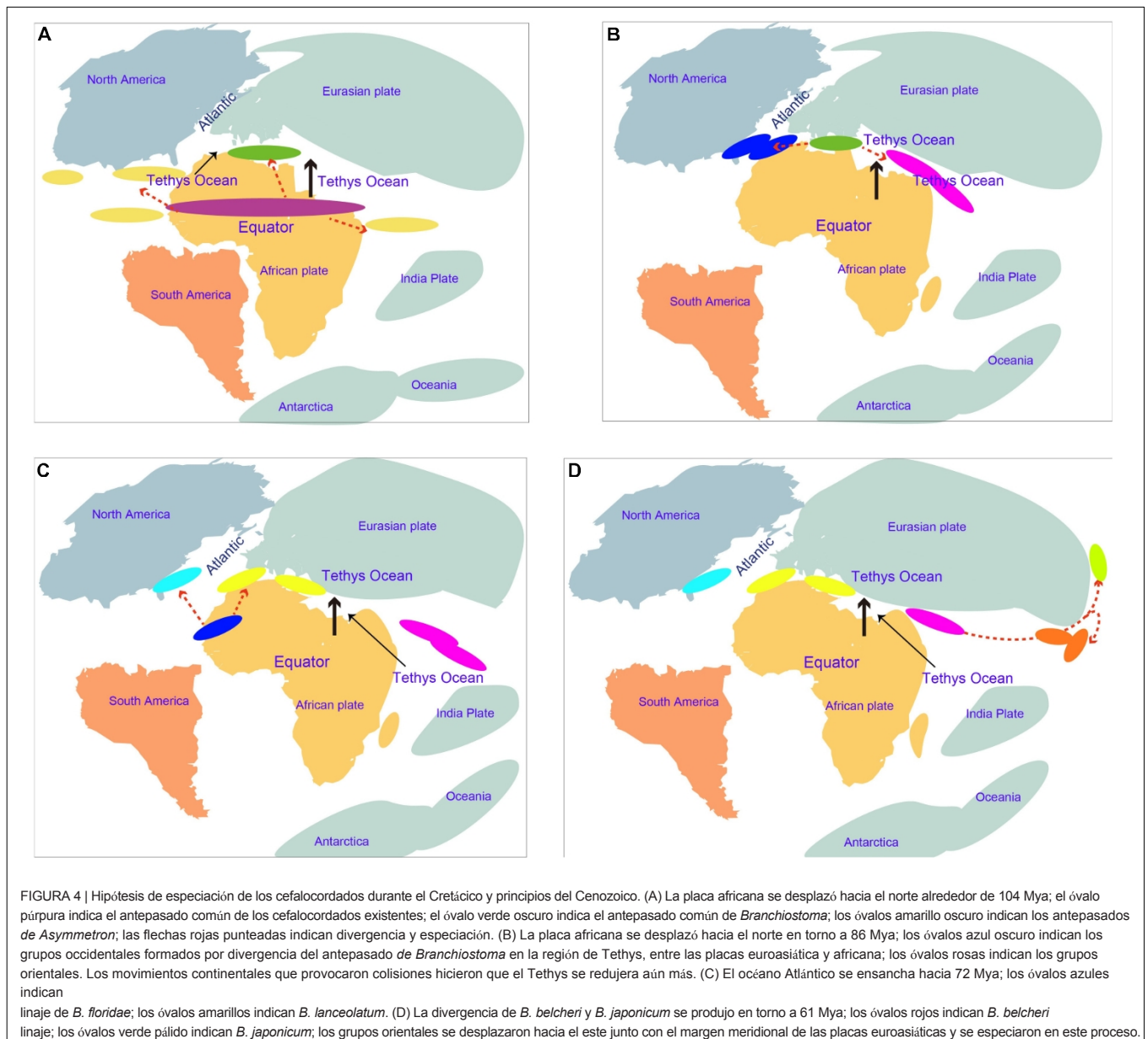
indicaron que las filogenias obtenidas a partir de la transcriptómica, genómico, o ADNmt completo representan el mayor rendimiento filogenético y permiten obtener mejores resultados que las filogenias a partir de conjuntos de genes únicos y a pequeña escala (Philippe et al., 2005; Yuan et al., 2018). Por lo tanto, los resultados obtenidos indicaron

B. lanceolatum como grupo hermano de *B. floridae*, en lugar de (*B. belcheri*+ *B. japonicum*). Estos análisis indicaron además que el ARNr 12S podía generar una filogenia similar a la de los genes nucleares masivos y el mitogenoma completo, por lo que puede utilizarse como posible marcador molecular en la investigación de la filogenia de los anfioxos. Sin duda, esta sugerencia debe investigarse más a fondo con más especies de amphioxus. Además, Yue et al. (2014) informaron de una tasa evolutiva más lenta de *B. floridae* y

A. lucayanum, incluso en comparación con el vertebrado de evolución más lenta conocido (el tiburón elefante), y los autores especularon con la posibilidad de también los géneros *Asymmetron* y *Branchiostoma*. De hecho, las especies de anfioxo *B. belcheri* y

B. japonicum también mostró un ritmo evolutivo extremadamente lento, apoyando la especulación de Yue et al. (2014). Sin embargo, descubrimos que *B. lanceolatum* evoluciona rápidamente, incluso por encima del nivel de los vertebrados de evolución más rápida. *B. lanceolatum* (denominado anfioxo mediterráneo) se encuentra principalmente en la costa del mayor mar intercontinental (el mar Mediterráneo) y en un mar continental (el mar Negro). Debido a estos entornos relativamente cerrados, a su descarga terrestre y a su clima único, ambas regiones (especialmente el Mediterráneo) muestran una diversidad biológica excepcionalmente alta (Meynard et al., 2012). Por lo tanto, un hábitat relativamente cerrado puede ser un factor clave para impulsar la rápida evolución de *B. lanceolatum* impulsada por su menor capacidad para amortiguar los factores ecológicos cambiantes. Se ha informado de que el agua de mar, calentada por una central nuclear desde 1980, causa una rápida evolución de la resistencia del parásito en la perca europea del casi cerrado mar Báltico (Mateosgonzalez et al., 2015). Además, una de las ramas más largas del árbol filogenético es la que conecta el anfioxo con los demás clados. Cerca de esta rama se encuentran varias de las otras ramas más largas de la filogenia. Este patrón parece estar causado por la atracción de ramas largas (LBA). No obstante, para la construcción del árbol filogenético se emplearon cuatro métodos utilizados para el establecimiento de árboles de genes y especies. En particular, el método de análisis ML presentó una posibilidad relativamente menor de generar LBA (Li et al., 2007). En cuanto al árbol de consenso posterior, debido a la implementación bayesiana de modelos de codones heterogéneos, CAT-GTR en PhyloBayes parece ser significativamente más robusto frente a LBA, en comparación con todos los demás modelos (Lartillot et al., 2013). Además, la posición filogenética de las ramas más largas fue en su mayoría coherente con lo informado anteriormente (Huang et al., 2014; Yue et al., 2014). Así pues, nuestro análisis ha evitado en la medida lo posible el LBA. Sin duda, requiere un análisis más exhaustivo para reducir aún más la posibilidad de LBA en futuros estudios.

La estimación del tiempo de divergencia es muy útil para comprender el evento de especiación y la historia evolutiva de



amphioxus. Los resultados de datación molecular obtenidos pueden estar sustancialmente influenciados por una serie de factores, por ejemplo, modelos evolutivos, métodos de análisis y limitaciones de tiempo (Baele et al., 2012; Yuan et al., 2018). Este estudio utilizó múltiples métodos para estimar el tiempo de divergencia de amphioxus para evitar el sesgo de usar un solo método de análisis. Además, se utilizaron muchas calibraciones de fósiles (alcance 10) y una restricción secundaria para mejorar la precisión lograda de la datación molecular. Nuestra estimación de la edad para los principales grupos de vertebrados (ligeramente más antigua) y la divergencia de *Branchiostoma-Asymmetron* (ligeramente más joven) fue en su mayor parte coherente con la comunicada por Yue et al. (2014), lo que indica que nuestros resultados obtenidos fueron sólidos. Cinco especies de anfioxos mostraron una especiación de no más de ~ 47 Mya (nodos p a m), lo que indica que aunque la mayoría de las especies de anfioxos mostraron un ritmo evolutivo lento (Putnam et al., 2008;

Yue et al., 2014), su proceso de especiación no es tan lento como cabría esperar, e incluso supera al de la mayoría de los grupos de vertebrados. A pesar de que el transcriptoma de *Epigonichthys*, el tercer género de cefalocordados, sigue ausente, el análisis del ADNmt completo mostró una relación filogenética de {vertebrados + [*Asymmetron* + (*Branchiostoma* + *Epigonichthys*)]}, lo que demuestra que *Asymmetron* divergió antes que los otros dos géneros de cefalocordados (Kon et al., 2007). Si se añadieran secuencias masivas de *Epigonichthys* a nuestros análisis, el primer tiempo de divergencia (~ 104 Mya) de los cefalocordados vivos no sería anterior al resultado actual. Antes y en esta época, expansión de la dorsal oceánica media impulsaba hacia el norte la placa africana según la hipótesis de Wegener de la deriva continental (Smith, 2014). El movimiento dividió a los antepasados comunes de *Asymmetron* y *Branchiostoma* en múltiples

poblaciones geográficas aisladas. Parte de ellas permanecieron en sus zonas de origen a lo largo del ecuador y evolucionaron hacia el ancestro *Asymmetron*, mientras que otras fueron empujadas hacia una altitud mayor y evolucionaron hacia el ancestro *Branchiostoma* en la región de Tethys, entre las placas euroasiática y africana (**Figura 4A**). Los *Asymmetron* se distribuyeron principalmente a lo largo del ecuador y latitudes bajas, mostrando una gran similitud con su patrón de distribución actual (Kon et al., 2006). Esto apoya aún más la hipótesis de divergencia propuesta entre *Asymmetron* y *Branchiostoma*. Además, es razonable especular que el ancestro común de los cefalocordados vivientes probablemente habitó zonas a lo largo del ecuador y latitudes bajas. Con el movimiento hacia el norte de la placa africana hacia la placa euroasiática, los antepasados de *Branchiostoma* se dividieron en grupos orientales y occidentales en esta región de Tethys (**Figura 4B**). En particular, debido a la fiable comunicación génica provocada por el cierre en gran medida incompleto del Tethys durante este periodo, sugerimos que la especiación entre ambos grupos permaneció inacabada hasta ~ 87 Mya. Alrededor de 110-65 Myr, el océano Atlántico se abrió gradualmente debido a la separación paulatina a mediados del Cretácico de Norteamérica de la placa euroasiática y a la ruptura de Gondwana (es decir, África y). La extensión del océano Atlántico constituye una barrera oceánica debido a la enorme profundidad y anchura entre Europa y América (Frisch y Dawes, 2014; Hou y Li, 2017). Por lo tanto, es probable que los grupos occidentales divergieran en *B. floridae* en la costa occidental y *B. lanceolatum* en la costa oriental del océano Atlántico en ~ 72 Mya debido a la expansión del océano Atlántico (**Figura 4C**). Se han notificado sugerencias similares para moluscos y crustáceos (Hou y Li, 2017). La India se desplazó hacia el norte desde su posición original y colisionó con Asia durante el Eoceno temprano (~ 50 Mya) (Hou y Li, 2017). De este modo, quedó por abrir una amplia región oriental del Tethys

~61 Mya (Hou y Li, 2017), y no estaba dividido en Océano Índico y Océano Pacífico. Por lo tanto, especulamos que los grupos orientales originales de la región de Tethys, entre las placas euroasiática y africana, se desplazaron hacia el este a través de la región entre las placas euroasiática e india. Posteriormente, parte de esta población en movimiento probablemente se trasladó a latitudes más altas a lo largo de la costa oriental de la placa euroasiática, y la especiación de *B. belcheri* y *B. japonicum* se produjo en ~ 61 Mya durante este periodo (**Figura 4D**). Es probable que la divergencia de ambas especies se viera impulsada por los cambios de hábitat causados por esta diferencia de latitud. Así pues, proponemos que la evolución oceánica impulsada por el movimiento platónico proporcionó un factor importante para la divergencia y especiación de los cefalocordados. En consecuencia, el patrón de distribución de los cefalocordados vivos guarda una estrecha relación evolutiva con la estructura tectónica.

El análisis comparativo de dominios entre diferentes especies de amphioxus no mostró diferencias significativas del número de dominios identificados por los genomas completos y los transcriptomas. Esto indica que los tipos de dominios (no la cantidad) de las especies se obtuvieron completamente a través del transcriptoma de los anfioxos adultos. A través de los análisis de enriquecimiento GO, encontramos que los dominios específicos de las especies de amphioxus están consistentemente relacionados con la respuesta inmune y la apoptosis. Este resultado indica que la exposición

a los agentes patógenos del agua de mar como fuerza impulsora principal para el origen y la evolución de los dominios de los anfioxos. Este

puede verse favorecida por su enterramiento en la arena del fondo marino y su relativamente escasa capacidad natatoria, lo que conduce a una acumulación de microorganismos patógenos. Estudios anteriores informaron de que los genes que evolucionan rápidamente entre diferentes géneros o especies de anfioxos estaban implicados principalmente en la inmunidad innata (Yue et Zhang et al., 2014; , 2018). Esto sugiere que la rápida evolución en la secuencia desempeñó un papel clave en el origen evolutivo de los dominios específicos de cada especie. Varias especies de amphioxus poseen capacidades especiales para la utilización de lípidos, en particular³

B. lanceolatum (el sitio web del proyecto de secuenciación del genoma de *B. lanceolatum*). Nuestras investigaciones hallaron varios términos GO relacionados con el almacenamiento, el metabolismo y la regulación de lípidos en

B. lanceolatum y *B. japonicum*, tal vez indicando que ambas especies son especies modelo ideales para la investigación de la biología lipídica de los amphioxus. Además, este resultado también indica que el ajuste específico de cada especie en la utilización y producción de energía es un paso evolutivo importante en la historia de especiación de los anfioxos. Los dominios primitivos comunes de los cefalocordados participaron principalmente en el desarrollo, el proceso y la función celular y el proceso metabólico para garantizar los procesos biológicos básicos normales.

En resumen, este estudio representa el primer análisis filogenómico a gran escala que incluye la mayoría de los principales géneros de cefalocordados basado en datos transcriptómicos. Una filogenia de [(*B. belcheri*+ *B. japonicum*)+ (*B. lanceolatum*+ *B. floridae*+ *A. lucayanum*)], aclarando la posición filogenética de *B. lanceolatum*. Además, propusimos que los genes mitocondriales del ARN 12S descritos anteriormente pueden ser un marcador molecular fiable en los análisis filogenéticos de los cefalocordados. En primer lugar, propusimos que las especies de anfioxo actualmente vivas no eran un grupo antiguo, y que muy probablemente irradiaron durante el Cretácico, y su diversificación y especiación fue impulsada principalmente por los movimientos platónicos durante este periodo. Además, se identificó y analizó funcionalmente un lote de dominios proteicos específicos de cada especie y ancestrales entre los cefalocordados mediante bioinformática, lo que indica varios factores que probablemente promovieron la adaptación de las especies de amphioxus a sus respectivos hábitats. Nuestro estudio se suma a la comprensión de las relaciones evolutivas de los cefalocordados y proporciona una visión de su historia de divergencia y características de especiación.

CONTRIBUCIONES DE LOS AUTORES

L-BL, J-YC y Q-LZ concibieron y diseñaron los experimentos. Q-LZ, G-LZ, M-LY, Z-XD y H-WL realizaron los experimentos. Q-LZ, M-LY, JG, FW y X-YD analizaron los datos. Q-LZ, L-BL y J-YC escribieron el artículo. L-BL, M-LY y revisaron el manuscrito.

FINANCIACIÓN

Esta investigación ha contado con el apoyo de la Fundación Nacional de Ciencias Naturales de China (31760042 y 41272008) y del programa

³ <http://www.genoscope.cns.fr/spip/L-espece-d-amphioxus-un-modele.html>

B para doctorandos destacados de la Universidad de Nanjing (subvención n.º 201702B075).

AGRADECIMIENTOS

Damos las gracias al profesor Shi-Cui Zhang de la Universidad Oceánica de China por ofrecernos materiales de *B. japonicum* amphioxus.

MATERIAL SUPLEMENTARIO

El material suplementario de este artículo puede consultarse en línea en: <https://www.frontiersin.org/articles/10.3389/fphys.2018.01833/full#supplementary-material>

FIGURA S1| Árbol filogenético bayesiano de 19 ramas inferido a partir de una matriz concatenada de genes ortólogos. Los números de las ramas representan las probabilidades bayesianas a posteriori.

FIGURA S2| Árbol filogenético PhyloBayes de 19 vías inferido a partir de una matriz concatenada de genes ortólogos.

FIGURA S3| Diagrama de Venn de dominios de cinco especies de amphioxus.

TABLA S1 | Resumen estadístico de la corrección de secuencias.

TABLA S2| Lista de términos GO enriquecidos por genes que contienen dominios compartidos entre todos los anfioxos analizados.

TABLA S3 | Información de dominios específicos de linaje en *Branchiostoma belcheri*.

TABLA S4 | Información de dominios específicos de linaje en *Branchiostoma floridae*.

TABLA S5| Información de dominios específicos de linaje en *Branchiostoma lanceolatum*.

TABLA S6 | Información de dominios específicos de linaje en *Branchiostoma japonicum*.

TABLA S7 | Información de dominios específicos de linaje en *Asymmetron lucayanum*.

TABLA S8| Lista de términos GO enriquecidos por genes que contienen dominios específicos de linaje en cada especie de amphioxus.

REFERENCIAS

- Abascal, F., Zardoya, R., y Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104-2105. doi: 10.1093/bioinformatics/bti263
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., y Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* 29, 2157-2167. doi: 10.1093/molbev/mss084
- Benton, M. J., y Donoghue, P. C. (2007). Pruebas paleontológicas para datar el árbol de la vida. *Mol. Biol. Evol.* 24, 26-53. doi: 10.1093/molbev/msl150
- Brown, J., Pirrung, M., y Mccue, L. A. (2017). FQC dashboard: integrate FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33, 3137-3139. doi: 10.1093/bioinformatics/btx373
- Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazonvega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Gene Dev.* 25, 1915-1927. doi: 10.1101/gad.17446611
- Chen, J. Y., Huang, D. Y., y Li, C. W. (1999). An early Cambrian crania te-like chordate. *Nature* 402, 518-522. doi: 10.1038/990080
- Conesa, A., Götz, S., García Gómez, J. M., Terol, J., Talón, M., y Roble s, M. (2005). Blast2GO: una herramienta universal para la anotación, visualización y análisis en la investigación genómica funcional. *Bioinformatics* 21, 3674-3676. doi: 10.1093/bioinformatics/bti610
- Dos, R. M., y Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* 28, 2161-2172. doi: 10.1093/molbev/msr045
- Ebersberger, I., Strauss, S., y von Haeseler, A. (2009). HaMStR: perfil modelo de markov oculto basado en la búsqueda de ortólogos en ESTs. *BMC Evol. Biol.* 9:157. doi: 10.1186/1471-2148-9-157
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy cy and high throughput. *Nucleic Acids Res.* 32, 1792-1797. doi: 10.1093/nar/gkh340
- Finn, R. D., Clements, J., y Eddy, S. R. (2011). Servidor web HMMER: búsqueda interactiva de similitud de secuencias. *Nucleic Acids Res.* 39, W29-W37. doi: 10.1093/nar/gkr367
- Frisch, T., y Dawes, P. R. (2014). Las rotaciones que abren los océanos Atlántico Central y Atlántico Norte: compilación, líneas de deriva y líneas de flujo. *Int. J. Earth Sci.* 103, 967-969. doi: 10.1007/s00531-013-0980-7
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644-652. doi: 10.1038/nbt.1883
- Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., y Gascuel, O. (2010). Nuevos algoritmos y métodos para estimar la máxima verosimilitud filogenias: evaluación del rendimiento de PhyML 3.0. *Syst. Biol.* 59, 307-321. doi: 10.1093/sysbio/syq010
- Guy, S. C. S., y Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6:31. doi: 10.1186/1471-2105-6-31
- Havird, J. C., y Santos, S. R. (2014). Performance of single and concatenated sets of mitochondrial genes at inferring metazoan relationships relative to full mitogenome data. *PLoS One* 9:e84080. doi: 10.1371/journal.pone.0084080
- Heled, J., y Drummond, A. J. (2009). Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* 27, 570-580. doi: 10.1093/molbev/ms p274
- Hou, Z., y Li, S. (2017). Tethyan changes shaped aquatic diversification. *Biol. Rev.* 93, 874-896. doi: 10.1111/brv.12376
- Huang, S., Chen, Z., Yan, X., Yu, T., Huang, G., Yan, Q., et al. (2014). Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat. Commun.* 5:5896. doi: 10.1038/ncomms6896
- Iseli, C., Jongeneel, C. V., y Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1999, 138-148.
- Itoh, M., Nacher, J. C., Kuma, K. I., Goto, S., y Kanehisa, M. (2007). Historia evolutiva e implicaciones funcionales de los dominios proteicos y sus combinaciones en eucariotas. *Genome Biol.* 8:R121. doi: 10.1186/gb-2007-8-6-r121
- Jantzen, S. G., Sutherland, B. J., Minkley, D. R., y Koop, B. F. (2011). GO trimming: systematically reducing redundancy in large gene ontology datasets. *BMC Res. Notes* 4:267. doi: 10.1186/1756-0500-4-267
- Kocot, K. M., Citarella, M. R., Moroz, L. L., y Halanych, K. M. (2013). PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinform. Online* 2013, 429-435. doi: 10.4137/EBO.S12813
- Kon, T., Nohara, M., Nishida, M., Sterrer, W., y Nishikawa, T. (2006). Hidden ancient diversification in the circumtropical lancelet *Asymmetron lucayanum* complex. *Mar. Biol.* 149, 875-883. doi: 10.1007/s00227-006-0271-y
- Kon, T., Nohara, M., Yamanoue, Y., Fujiwara, Y., Nishida, M., Nishikawa, M., et al. (2007). Phylogenetic position of a whale-fall lancelet (*Cephalochordata*) inferred from whole mitochondrial genome sequences. *BMC Evol. Biol.* 7:127. doi: 10.1186/1471-2148-7-127
- Koop, D., y Holland, L. Z. (2008). The basal chordate amphioxus as a simple model for elucidating developmental mechanisms in vertebrates. *Birth Defects Res. Part C Embryo Today Rev.* 84, 175-187. doi: 10.1002/bdrc.20128
- Kriventseva, E. V., Tegenfeldt, F., Petty, T. J., Waterhouse, R. M., Simão, F. A., Pozdnyakov, I. A., et al. (2015). OrthoDB v8: actualización del catálogo jerárquico de ortólogos y del software libre subyacente. *Nucleic Acids Res.* 43, D250- D256. doi: 10.1093/nar/gku1220

- Lartillot, N., Rodrigue, N., Stubbs, D., y Richer, J. (2013). PhyloBayes MPI: reconstrucción filogenética con infinitas mezclas de perfiles en un entorno paralelo. *Syst. Biol.* 62, 611-615. doi: 10.1093/sysbio/syt022
- Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., y Prohaska, S. J. (2011). Proteinortho: detección de (Co-)ortólogos en análisis a gran escala. *BMC Bioinformatics* 12:124. doi: 10.1186/1471-2105-12-124
- Li, F., Xiao, Y., Huang, F., Deng, W., Zhao, H., Shi, X., et al. (2015). Spatiotemporal-specific lncRNAs in the brain, colon, liver and lung of macaque during development. *Mol. Biosyst.* 11, 3253-3263. doi: 10.1039/c5mb00474h
- Li, Y. W., Yu, L., y Zhang, Y. P. (2007). "Long-branch Attraction" artifact in phylogenetic reconstruction. *Heredity* 29, 659-667. doi: 10.1360/yc-007-0659
- Löytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10557-10562. doi: 10.1073/pnas.0409137102
- Luo, J. C., Xiao, Y. S., Song, L., Gao, T. y Zhang, Y. (2007). Phylogenetic separation of lancelets in China revealed by mitochondrial COI and Cyt b gene analysis. *S. China Fish. Sci.* 3, 8-14.
- Mateosgonzalez, F., Sundström, L. F., Schmid, M., y Björklund, M. (2015). Rápida evolución de la resistencia de los parásitos en un ambiente más cálido: perspectivas de un experimento de campo a gran escala. *PLoS One* 10:e0128860. doi: 10.1371/journal.pone.0128860
- Meynard, C. N., Mouillot, D., Mouquet, N., y Douzery, E. J. P. (2012). A phylogenetic perspective on the evolution of mediterranean teleost fishes. *PLoS One* 7:e36443. doi: 10.1371/journal.pone.0036443
- Mirarab, S., Bayzid, M. S., y Warnow, T. (2016). Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Syst. Biol.* 65, 366-380. doi: 10.1093/sysbio/syu063
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., et al. (2014). La filogenómica resuelve el calendario y el patrón de la evolución de los insectos. *Science* 346, 763-767. doi: 10.1126/science.1257570
- Moreno-Hagsieb, G., y Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* 24, 319-324. doi: 10.1093/bioinformatics/btm585
- Mount, D. W. (2007). Uso de la herramienta básica de búsqueda de alineación local (BLAST). *CSH Protoc.* 2007.pdb.top17. doi: 10.1101/pdb.top17
- Mulcahy, D. G., Noonan, B. P., Moss, T., Townsend, T. M., Reeder, T. W., Sites, J. W., et al. (2012). Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Mol. Phylog. Evol.* 65, 974-991. doi: 10.1016/j.ympev.2012.08.018
- Peterson, K. J., Cotton, J. A., Gehling, J. G., y Pisani, D. (2008). The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1435-1443. doi: 10.1098/rstb.2007.2233
- Philippe, H., Delsuc, F., Brinkmann, H., y Lartillot, N. (2005). Phylogenomics. *Annu. Rev. Ecol. Syst.* 36, 541-562. doi: 10.1146/annurev.ecolsys.35.112202.130205
- Putnam, N. H., Butts, T., Ferrier, D. E. K., Furlong, R. F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064-1071. doi: 10.1038/nature06967
- Rannala, B., y Yang, Z. (2007). Inferring speciation times under an episodic molecular clock. *Syst. Biol.* 56, 453-466. doi: 10.1080/10635150701420643
- Rekapalli, B., Halloy, C., y Zhulin, I. B. (2009). "HSP-HMMER: a tool for protein domain identification on a large scale", en *Proceedings of the ACM Symposium on Applied Computing* (Nueva York, NY: ACM, Inc.), 766-770.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539-542. doi: 10.1093/sysbio/sys029
- Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101-109. doi: 10.1093/oxfordjournals.molbev.a003974
- Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301-302. doi: 10.1093/bioinformatics/19.2.301
- Sayyari, E., y Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33, 1654-1668. doi: 10.1093/molbev/msw079
- Schwartz, R. S., y Mueller, R. L. (2010). Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evol. Biol.* 10:5. doi: 10.1186/1471-2148-10-5
- Shimeld, S. M., y Holland, N. D. (2005). Amphioxus molecular biology: insights into vertebrate evolution and developmental mechanisms. *Rev. Canadienne De Zool.* 83, 90-100. doi: 10.1139/z04-155
- Silvan, O., Stephanie, B., Belgacem, M. R., Yann, L. P., y Hector, E. (2012). Secuenciación y análisis del transcriptoma del anfioxo mediterráneo (*Branchiostoma lanceolatum*). *PLoS One* 7:e36554. doi: 10.1371/journal.pone.0036554
- Smith, H. A. (2014). Las pruebas geodésicas relativas a la hipótesis de Wegener. *Empire Surv. Rev.* 9, 90-100. doi: 10.1179/sre.1947.9.65.90
- Sun, Y. B. (2017). FasParser: un paquete para manipular datos de secuencias. *Zool. Res.* 38, 110-112. doi: 10.24272/j.issn.2095-8137.2017.017
- Tajima, F. (1993). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135, 599-607.
- Talavera, G., y Castresana, J. (2007). Mejora de las filogenias tras eliminar los bloques divergentes y ambiguamente alineados de los alineamientos de secuencias proteicas. *Syst. Biol.* 56, 564-577. doi: 10.1080/10635150701472164
- Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., et al. (2014). Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505, 174-179. doi: 10.1038/nature12826
- Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., y Teichmann, S. A. (2004). Estructura, función y evolución de las proteínas multidominio. *Curr. Opin. Struct. Biol.* 14, 208-216. doi: 10.1016/j.sbi.2004.03.011
- Yuan, M. L., Zhang, Q. L., Guo, Z. L., Wang, J., y Shen, Y. Y. (2015). El genoma mitocondrial completo de *Corizus tetraspilus* (Hemiptera: Rhopalidae) y análisis filogenético de Pentatomomorpha. *PLoS One* 10:e0129003. doi: 10.1371/journal.pone.0129003
- Yuan, M. L., Zhang, Q. L., Zhang, L., Guo, Z. L., Liu, Y. J., Shen, Y. Y., et al. (2016). Filogenia de alto nivel de los Coleoptera inferida con secuencias del genoma mitocondrial. *Mol. Phylog. Evol.* 104, 99-111. doi: 10.1016/j.ympev.2016.08.002
- Yuan, M. L., Zhang, Q. L., Jia, C. L., Li, X. P., Yang, X. Z., et al. (2018). Filogenia mitocondrial, historia de divergencia y adaptación a gran altitud de orugas de pastizales (Lepidoptera: Lymantriinae: Gynaephora) que habitan en la meseta tibetana. *Mol. Phylog. Evol.* 122, 116-124. doi: 10.1016/j.ympev.2018.01.016
- Yue, J. X., Yu, J. K., Putnam, N. H., y Holland, L. Z. (2014). El transcriptoma de un anfioxo, *Asymmetron lucayanum*, de las Bahamas: una ventana a la evolución de los cordados. *Genome Biol. Evol.* 6, 2681-2696. doi: 10.1093/gbe/evu212
- Zhang, Q. L., Xu, B., Wang, X. Q., Yuan, M. L., y Chen, J. Y. (2018). Genome-wide comparison of the protein-coding repertoire reveals fast evolution of immune-related genes in cephalochordates and *Osteichthyes superclass*. *Oncotarget* 9:83. doi: 10.18632/oncotarget.22749

Declaración de conflicto de intereses: Los autores declaran que la investigación se llevó a cabo en ausencia de relaciones comerciales o financieras que interpretarse como un posible conflicto de intereses.

Copyright © 2018 Zhang, Zhang, Yuan, Dong, Li, Guo, Wang, Deng, Chen y Lin. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia de Atribución Creative Commons (CC BY). Se su uso, distribución o reproducción en otros foros, siempre que se cite al autor o autores originales y al propietario o propietarios de los derechos de autor y que se cite la publicación original en esta revista, de acuerdo con la práctica académica aceptada. No se permite ningún uso, distribución o reproducción que no cumpla estas condiciones.